Multivariate Distribution

- Let
$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$
random vector consisting of p random variables.

- The cumulative distribution function (CDF) of X is defined by
$$F(\mathbf{x}) = P(X \leq \mathbf{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p).$$

- The marginal cdf: $F_{X_1}(x_1) = P(X_1 \leq x_1) = F(x_1, \infty, \ldots, \infty)$

- For continuous X, density f exists such that
$$\begin{aligned} F(x) &= \int_{-\infty}^{\mathbf{x}} f(\mathbf{u}) d\mathbf{u} \\ &= \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(u_1, \ldots, u_p) du_1 \ldots du_p \end{aligned}$$

Moments

For random vectors $X_{p\times 1}$ and $Y_{q\times 1}$,

- (population) mean vector:

$$\mu_X = \begin{pmatrix} EX_1 \\ EX_2 \\ \vdots \\ EX_p \end{pmatrix}$$

- (population) variance-covariance matrix:

$$\begin{aligned} Var(X) = \Sigma_X &= E((X - EX)(X - EX)^T) \\ &= E\left(\begin{pmatrix} X_1 - EX_1 \\ X_2 - EX_2 \\ \vdots \\ X_p - EX_p \end{pmatrix} \begin{pmatrix} X_1 - EX_1 & X_2 - EX_2 & \cdots X_p - EX_p \end{pmatrix} \right) \end{aligned}$$

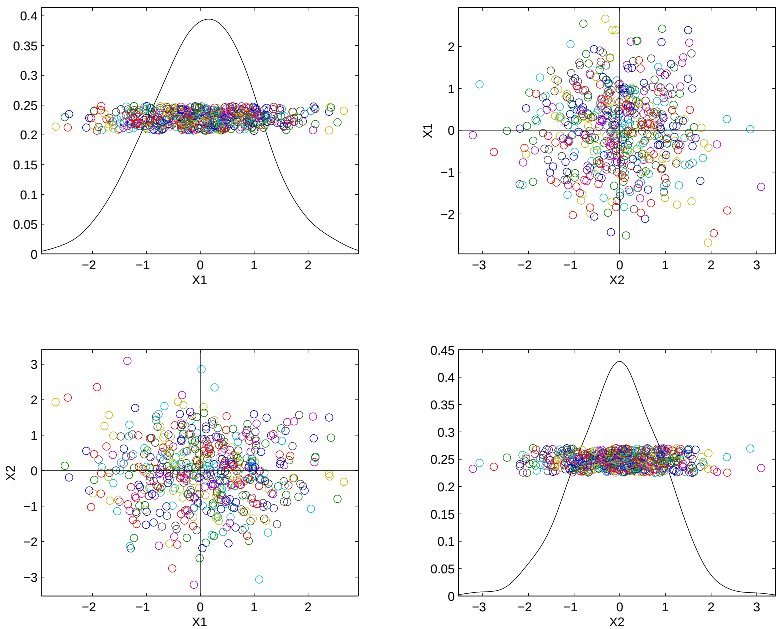
- (population) covariance matrix:

$$\Sigma_{XY} = Cov(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)^T\}.$$

Properties of Means, Variances and Covariances

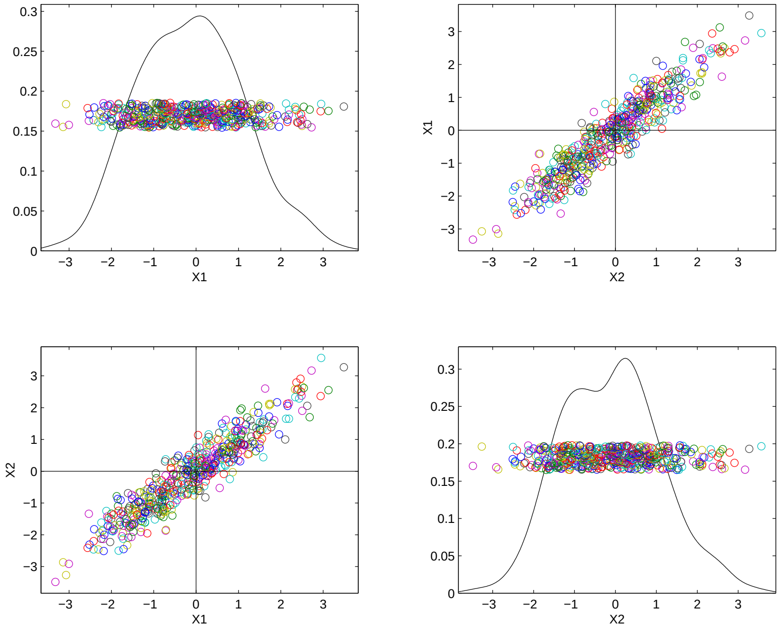
- $E(AX) = AE(X)$
- $\Sigma = E(XX^T) - \mu\mu^T$
- $\Sigma \geq 0$
- $Var(AX + b) = AVar(X)A^T$
- $Var(\mathbf{c}^T X) = \mathbf{c}^T Var(X)\mathbf{c}$
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
- $Cov(AX, BY) = ACov(X, Y)B^T$

Example1: Scatter Matrix plot



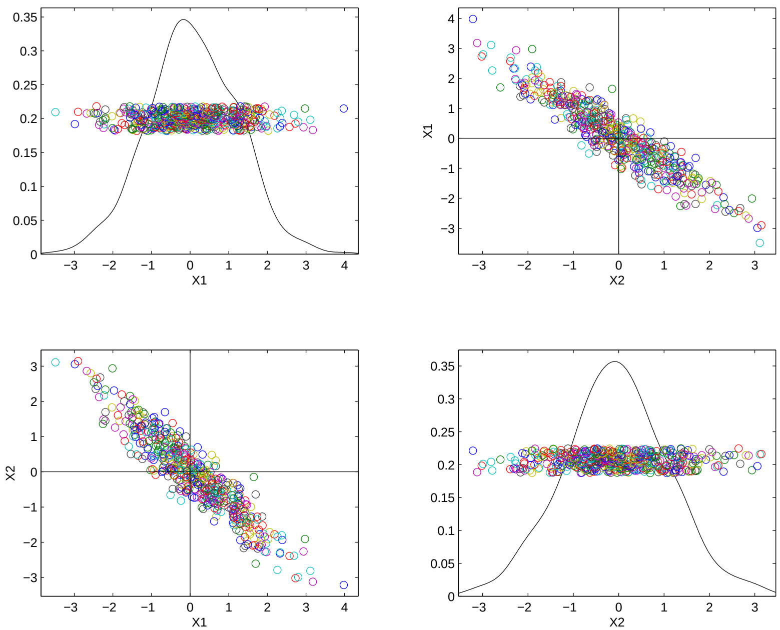
$$X \sim N_2(0, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})$$

Example1: Scatter Matrix plot



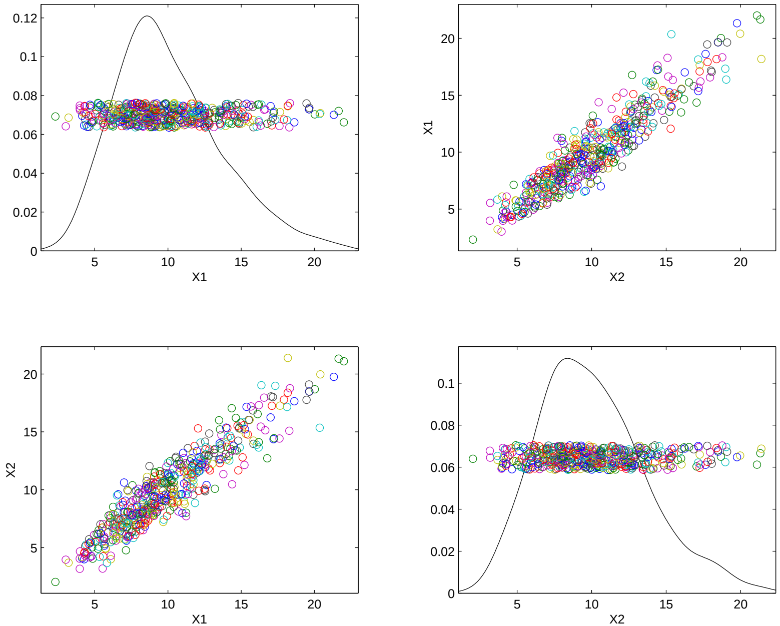
$$X \sim N_2(0, \begin{pmatrix} 1 & .7 \\ .7 & 1 \end{pmatrix})$$

Example1: Scatter Matrix plot



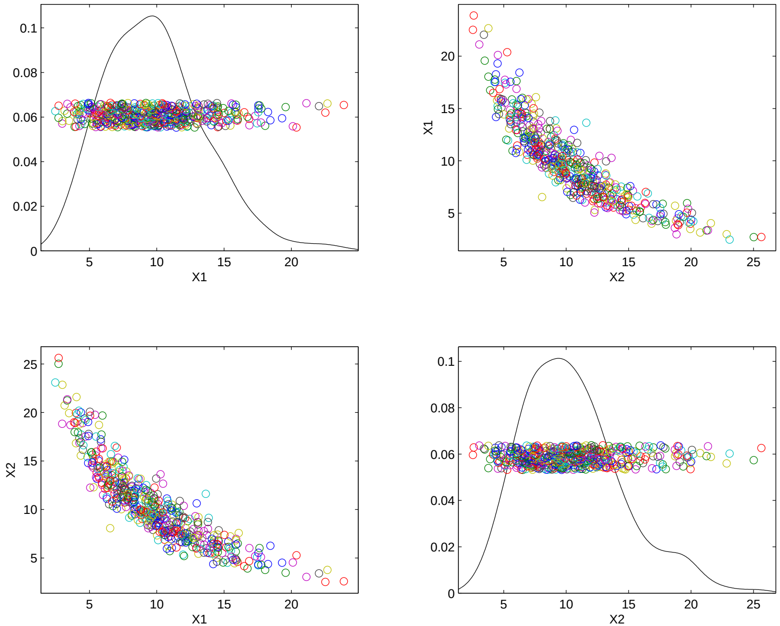
$$X \sim N_2(0, \begin{pmatrix} 1 & -.7 \\ -.7 & 1 \end{pmatrix})$$

Example2: Scatter Matrix plot



- $X \sim N_2(0, \begin{pmatrix} 1 & .7 \\ .7 & 1 \end{pmatrix})$
- $(U_1, U_2) = (F_1(X_1), F_2(X_2))$ for $i = 1, 2$ where F_is are the marginal cdf of X.
- $(Y_1, Y_2) = (F_G^{-1}(U_1), F_G^{-1}(U_2))$, where F_G^{-1} is the inverse of the Gamma $(10, 1)$ cdf.

Example2: Scatter Matrix plot



- $X \sim N_2(0, \begin{pmatrix} 1 & -.7 \\ -.7 & 1 \end{pmatrix})$
- $(U_1, U_2) = (F_1(X_1), F_2(X_2))$ for $i = 1, 2$ where F_is are the marginal cdf of X.
- $(Y_1, Y_2) = (F_G^{-1}(U_1), F_G^{-1}(U_2))$, where F_G^{-1} is the inverse of the Gamma $(10, 1)$ cdf.

Random sample

- The random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are said to form a random sample of size n from $F(\mathbf{x})$ if
  - $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are mutually independent and
  - the distribution of each $\mathbf{X}_i$ is $F(\mathbf{x})$.
- For a random sample of $\mathbf{X}_1, \ldots, \mathbf{X}_n$, the data matrix is defined as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \vdots \\ \mathbf{X}_n^T \end{bmatrix}.$$

- Note that $\mathbf{X}$ is a random matrix where the row of the matrix are independent, but the columns may may not.

Sampling distributions

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from a distribution that has mean vector μ and covariance matrix Σ. Then, the mean vector

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i$$

and the covariance matrix

$$\mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$$

satisfies the following:

- $E\bar{\mathbf{X}} = \mu$
- $Cov(\bar{\mathbf{X}}) = \frac{1}{n}\Sigma$
- $E\mathbf{S}_n = \Sigma.$

Linear combination of random variables

In many multivariate procedures, we will look at a linear combinations of the form

$$\mathbf{c}^T X = c_1 X_1 + c_2 X_2 + \ldots + c_p X_p,$$

where $\mathbf{c} = (c_1, \ldots, c_p)^T \in \mathbb{R}^p$. Let $Y = \mathbf{c}^T X$ be the transformed random variable.

- What is $E(Y)$?
- What is $\mathrm{Var}(Y)$?

Given a random sample of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ for X, what is the random sample of Y?

- Sample mean?
- Sample variance?